

# CEI: a new indicator measuring City Commercial Credit Risk initiated in China

Ruonan Lin<sup>1\*</sup> and Yi Gu<sup>1,2</sup>

<sup>1</sup>*Department of Applied Mathematics and Statistics,  
Stony Brook University, Stony Brook, NY 11794, USA*

<sup>2</sup>*Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794, USA*

(Dated: May 20, 2015)

Aiming at quantifying and evaluating the regional commercial environment along with the level of economic development among cities in mainland China, the concept of China City Commercial Environment Credit Index(CEI) was first introduced and established in 2010. In this manuscript, a historical review and detailed introduction of CEI is included, followed by statistical studies. In particular, an independent statistical cross-check for the existing CEI-2012 is performed and significant factors that play the most influential roles are discussed.

## I. HISTORY REVIEW AND PAPER STRUCTURE

China, one of the fastest growing economies in the world, has always been attracting worldwide attention in the past two decades. After a period of stable and rapid development, China has entered a new era of credit economy. By the end of last century, China initiated to build its own social credit system, of which the scale and importance are almost comparable to the social security system in the United States. In just over one decade of time, this multi-facet system, from its coming into existence, has already exhibited increasing power of steering China's economic and social development.

Social credit system functions to increase credit launch, control credit risk and enhance credit awareness[1]. Chinese government expects the credit system to act in avoiding, controlling and transferring credit risk as well as to establish a "credit punishment mechanism", based upon which it is also anticipated to help regulate market behaviors and nurture new business philosophy. Because of the "punishment mechanism", social credit system becomes the only feasible solution to dishonest deed and discreditable behaviors, and thus it is crucial to the construction and regulation of market within framework of credit economy.

Not until recently did China realize that its government is in urgent need of creating a series of economic indices to help evaluate the development level of urban social credit system and to quantify the commercial credit risk. To this end, China City Commercial Environment Credit Index(China City CEI, or CEI for short) was introduced by Chinese Academy of Management Science in 2010. Each participating city is assigned a numerical value(within 30-100), which is carefully calculated via a specific statistical model. Taking into account multiple commercial factors, the CEI is expected to reflect the status quo of local credit system, and it serves as a tool to evaluate commercial environment. Since "city" is the unit entity in the system, CEI is calculated on a city-wide

basis. As an index computed and tested annually, CEI helps to build the whole social credit system in that it quantifies all the relevant commercial factors in the realm of economics and sociology, and it makes the credit system more concrete, measureable and comparable.

A group of CEI initiators has tested this economic index among 248 cities in China, and researchers have been working diligently on its improvement. After the first two release periods(CEI for year 2011, 2012), both the central and local government in China have paid great attention to it. The CEI will surely influence the government policy-making as one of the most important economic indices to refer to in future.

In this paper we include a relatively detailed introduction of CEI in the second part, including factors involved that are important to construct CEI. In the third part, on top of the existing categorical CEI-2012(i.e., the CEI's based on each index from the first index class, see later section), we present an independent statistical cross-check using Principal Component Analysis(PCA) techniques, from which the most influential factors in CEI are extracted. A paired test is also performed, and the result from the test indicates that CEI is indeed a convincing tool to evaluate the status quo of commercial environment among cities in mainland China. In the last part of the paper, the importance and future influence of CEI will be discussed.

## II. INTRODUCTION

CEI, an unprecedented example of applying theories of social credit system into practice, has been regarded as a remarkable milestone over the course of city development history in China. As a comprehensive economic index, CEI is being used to measure and evaluate the level of credit growth, the trading levels in credit market, the goodness of a market, and the overall level of social credit awareness. It enables us to analyze the complicated social economics phenomena. To be more specific, CEI reflects, but not limited to, the following regional feature of commercial environment [2]:

\* Correspondence to: linruonan@gmail.com

• The activities of urban credit economics, for exam-

ple, the variation of credit instrument launch and enterprise spot sale on credit

- The fluctuation and operational level of urban credit economics
- The scale of urban credit transactions
- The risk level of credit transactions in local market
- The establishment of risk by-law among local enterprises
- The precautions steps that local enterprises take to avoid, control and transfer credit risk
- To what extent the local market environment influences local enterprise's good standing and development

### A. Evaluation Index System

Given the resources we have for collecting reliable and utilizable credit information(data) in China, CEI is established upon the following multiple commercial factors, namely credit launch, enterprise risk management, construction of credit reporting system, credit monitoring services from local government, the odds of dishonesty deed and discreditable behaviors, the build-up of credit reward and punishment system. These factors comprise the *first-class index* in the China city commercial environment evaluation index system.

Most of the indices in first-class reflect sociological aspects of group behavior, and therefore they are quite abstract and hard to directly quantify. For this reason we introduce *second-class index*, which are slightly more concrete. The *third-class index* is more or less the embodiment or concretization of the first two class indices. All the three class indices comprise the entire China city commercial environment evaluation index system. The entire index system is included in Table.I[3].

### B. Data Sources and Data Processing

CEI is a retrospective study of city credit environment, that is, it is calculated based upon data collected in the prior fiscal year. For example, the CEI released in July of 2012 utilizes the data collected between Apr.1<sup>st</sup>, 2011 and Mar.31<sup>st</sup>, 2012. By convention of economic index, CEI reflects the feature(commercial environment of a city) of last year. As information is collected every year, the cumulative data over a long period of time allow us to use regression analysis technique to predict future behaviors(i.e., city credit environment).

To ensure reliability of data and authority of CEI, credit information is collected simultaneously in all participating China cities according to stringent statistical model requirements. The CEI research group clearly defines

each index and its corresponding way of collecting and processing data. The standardized procedure of data-processing is universal among all participating cities so as to guarantee cross-city consistency and comparability.

The data source selection is also very cautious and demanding. For CEI-2012, the collected credit information only comes from the following sources: (1) major credit reporting agencies; (2) government statistics; (3) credit monitoring authorities; (4) indirect but authoritative data; for example, CEI research group utilizes Themis<sup>©</sup> financial risk early-warning index and its evaluation results to assess enterprise financial risk measures[4]; (5) questionnaires conducted by authorized research and development unit.

There are dozens of factors involved in modeling CEI(listed in Table.I), therefore CEI calculation requires a *comprehensive evaluation method*. This method uses multiple indices to evaluate multiple participating cities, and it is also named *multi-index comprehensive evaluation method*. The underlying idea is to transform multiple indices into one index that could reflect the comprehensive situation, and we use that index to evaluate each participating city.

Generally speaking, there are two major steps that are indispensable in the evaluation process[5]:

(a) quantify credit information and pretreat indices of different units so as to remove the influence of non-uniform dimension

(b) calculate comprehensive evaluation index(or comprehensive evaluation score) based on the post-treatment index values. If linear weighting method is applied to pick up contributions from all indices, a weighting factor should be assigned for each index in certain index class.

Using multi-objective programming principles and the efficacy coefficient method, one is able to find upper and lower limit of each single index, and we further transform each single index into comparable evaluation score, i.e., single index evaluation value. Let  $X_i$  denotes the raw value of index  $i$ , and  $X_{min}^i(X_{max}^i)$  be the minimum(maximum) raw value of index  $i$ .  $Z_i$  is the dimensionless evaluation score. For any "positive index"(the index whose higher value is favored), the usual efficacy coefficient can be calculated:

$$Z_i = \frac{x_i - x_{min}^i}{x_{max}^i - x_{min}^i} \times 100. \quad (1)$$

The modified efficacy coefficient can be found in Eq.(2)

$$Z_i = \frac{x_i - x_{min}^i}{x_{max}^i - x_{min}^i} \times 40 + 60. \quad (2)$$

After obtaining all dimensionless evaluation values(each corresponds to one index), the CEI research group uses linear weighted method to calculate overall

First-class Index	Second-class Index	Third-class Index
credit launch	credit instrument launch	the proportion of consumer credit balances in GDP
		the proportion of credit loan from small and medium-sized enterprises in overall credit loan
	commercial credit sales	the proportion of credit sales contracts in overall sales contracts
		the proportion of gross credit sales in GDP
		the proportion of agreement-honored credit sales in gross credit sales
enterprise risk management	establishment of credit management department in enterprise	the proportion of enterprise with credit management dept.
		the client enterprises of large credit reporting agencies
	risk management talents	the proportion of credit management licensee in all employees
		the employees that attended risk management training
		risk management major graduates from local institutes
	enterprise financial risk	Themis <sup>®</sup> financial risk early-warning index of the listed companies
construction of credit reporting system	development of credit-reporting industry	the variety/completeness of types of credit reporting agencies
		the number of credit reporting agencies
	credit info reports, related products	the variety of credit products and services provided
	utilization of public credit reporting system (on a yearly basis)	the proportion of enterprises whose credit info are consulted
credit monitoring from government	establishment of monitoring dept.	the proportion of individuals whose credit info are consulted
	credit infrastructure	local governments with credit monitoring department
		the well-being and upgrade status of credit infrastructure
	establishment and implementation of the by-law of credit transaction	whether “blacklist system” is set up in monitoring dept.
		the amount of by-law or regulations established
		the amount of credit standards established
discredit dishonesty deed	economic law cases	the implementation of national credit standards
		the number of economic crimes of swindling
		the amount of money involved in swindling crimes
	major discredit criminal cases	crimes of producing and selling fake and inferior goods
		major discredit criminal cases
credit education	the build-up of credit system	bribe and corruption crimes
	the status quo of credit education	whether personnel identified to take over the system build-up
		number of events on credit education
corporate experience	satisfaction with credit environment	whether credit education info available online(via web,vedio...)
		questionnaire that covers over 30 categories

TABLE I. China City Commercial Environment Evaluation Index System

CEI,

### III. STATISTICAL STUDIES

$$CEI = \frac{\sum_{i=1} Z_i W_i}{\sum_{i=1} W_i}. \quad (3)$$

where in Eq.(3),  $W_i$  is the weight for index  $i$ . Similar to the group decision making process, these weights are determined via Analytic Hierarchy Process(AHP) in subjective weighting method to minimize the influence of subjectivity.

We perform a set of completely independent statistical studies as a cross-check to the CEI released in year 2012 using Principal Component Analysis(PCA) technique.

PCA is a statistical method that uses orthogonal transformations to convert a set of possibly correlated variables into a set of linearly uncorrelated variables. These uncorrelated variables are called *principal components*. Such transformation is defined in a way that the first principal component accounts for as much of the variability in data as possible and could be used as a comprehensive strength indicator. The dependent variable in PCA

	LCH	ETP	SYS	GOV	DIS	EDU
LCH	1.0000	0.1576	0.1933	0.1539	0.0849	0.1654
ETP	0.1576	1.0000	-0.0149	-0.1805	-0.1170	-0.2306
SYS	0.1933	-0.0149	1.0000	0.3502	0.2343	0.4114
GOV	0.1539	-0.1805	0.3502	1.0000	0.2306	0.4438
DIS	0.0849	-0.1170	0.2343	0.2306	1.0000	0.0847
EDU	0.1654	-0.2306	0.4114	0.4438	0.0847	1.0000

TABLE II. covariance matrix for the six observables(indices) in first class. LCH: credit launch, ETP: enterprise risk management, SYS: construction of credit reporting system, GOV: credit monitoring at government level, DIS: discredit/dishonesty and EDU: credit education status

is not an actual existing variable, but a variable defined by the analyst. In our studies, the original independent variables are the official categorical CEI's(or city rankings by categorical CEI) based on six of the indices from first class: credit launch, enterprise risk management, construction of credit reporting system, credit monitoring at government level, discredit/dishonesty, credit education. The dependent variable is the overall City Commercial Environment Credit Index (CEI).

	Eigenvalue	Difference	Proportion	Cumulative
1	2.05339008	0.86295359	0.3422	0.3422
2	1.19043649	0.25762628	0.1984	0.5406
3	0.93281021	0.19024314	0.1555	0.6961
4	0.74256708	0.13428583	0.1238	0.8199
5	0.60828125	0.13576637	0.1014	0.9212
6	0.47251488		0.0788	1.0000

TABLE III. eigenvalues of covariance matrix in Table.II. The  $i^{th}$  row corresponds to the  $i^{th}$  eigenvector(principal component). "Proportion" denotes the proportion(or percentage) of variance in data specific eigenvector could account for, and "Cumulative" denotes cumulative proportion of variance in data if all eigenvectors up to current one(including current one) are included.

	Y1	Y2	Y3	Y4	Y5	Y6
LCH	0.244189	0.642825	-0.054923	-0.699277	-0.151514	-0.110388
ETP	-0.210139	0.728475	0.044816	0.473807	0.333973	0.295171
SYS	0.496945	0.199181	0.008052	0.531121	-0.488440	-0.438913
GOV	0.526149	-0.075805	-0.070938	-0.000772	0.783788	-0.313151
DIS	0.310926	-0.048091	0.899320	-0.050786	-0.012501	0.299165
EDU	0.525120	-0.091473	-0.425559	0.042999	-0.111586	0.721443

TABLE IV. eigenvectors of covariance matrix in Table.II. Variable abbr.'s follow the same rule as in Table.II

The maximum number of principal components can be same as number of independent variables. The first principal component reflects the "overall strength of response". The remaining principal components could be automatically calculated, and they may not have practical meaning. Each succedent component is required to

have the highest possible variance under given constraint, which is orthogonal to the preceding components.

The structural formula for the principal component is expressed as:

$$Z = \sum_i^p a_i u_i. \quad (4)$$

where  $p$  is the number of original independent variables.  $a_i$  is the coefficient reflecting the degree of influence that  $i^{th}$  variable casts on the principal component.  $u_i$  is the standardized value of variable  $i$ .

The covariance matrix is calculated after each evaluation score is normalized because PCA is sensitive to the relative scale of the variables. The results can be found in Table.II. The eigenvalues and eigenvectors of the covariance matrix can be found in Table.III and IV respectively.

From Table.III, we notice the largest eigenvalue 2.0534 corresponds to the first principal component(Y1). The eigenvector Y1 returns coefficients for the involved terms in the first principal component. Therefore the evaluation score based on the first principal component(denoted as Z1) can be expressed in Eq.(5):

$$\begin{aligned} Z1 = & 0.244189 * LCH - 0.210139 * ETP \\ & + 0.496945 * SYS + 0.526149 * GOV \\ & + 0.310926 * DIS + 0.525120 * EDU. \end{aligned} \quad (5)$$

where LCH(our original variable) is the official city rankings by categorical CEI based on the factor of credit launch. Similarly for others.

The expressions for other principal components in terms of six index values can be easily achieved in the same manner. The principal component score( $Z_i$ ) can be calculated accordingly.

One possible criterion for determining whether to keep or discard a component is the proportion(or percentage) of variance in data such component accounts for. For example, you may decide to keep any component that accounts for at least 15% or 20% of the total variance in your model and drop the others. The proportion is defined and calculated using the following simple formula:

$$\text{proportion} = \frac{\text{eigenvalue for the component of interest}}{\text{total eigenvalues of the correlation matrix}} \quad (6)$$

In principal component analysis, the "total eigenvalues of the correlation matrix" is equal to the total number of variables being analyzed, because each variable contributes one unit of variance to the analysis. Therefore, the larger the eigenvalue of the component is, the greater the proportion of variance in data that component accounts for. For example, the first principal component, with eigenvalue of 2.0534, accounts for approximately  $2.0534/6 \sim 34\%$  of the information in data. The second principal component comes next, which accounts for  $\sim 19\%$  of variance in data.

We also plot the eigenvectors of all six indices for the first two principal components Y1 and Y2 in Fig.1. Since the first principal component weights almost twice as much as that of the second principal component, the most influential indices are determined by the coefficient of each piece in Y1. Within Y1, GOV(credit monitoring at government level) and EDU(credit education status) contribute roughly equal amount, hence they are the most influential indices in modeling overall CEI.

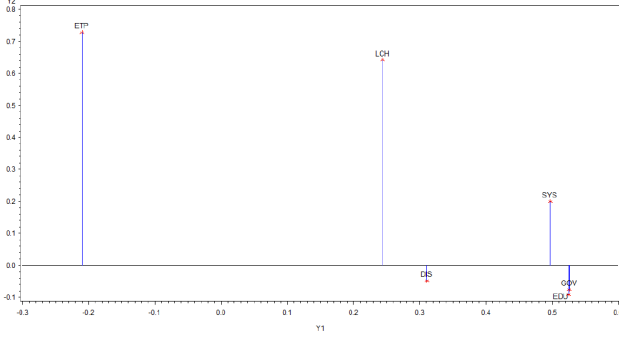


FIG. 1. Coefficients of the six indices involved in this study for the first two principal components Y1, Y2. The horizontal axis is for Y1, and vertical axis is for Y2.

A scatter plot of scores is also presented based upon the first principal component in Fig.2. To make our scores directly comparable to official overall CEI, we apply a simple transformation to scale up raw Z1 values into expected CEI range. We do not label all 284 cities in English because of the intrinsic nature of the data collected by our Chinese colleagues. The score is formatted so that higher value indicates better credit environment. For example, the city of Fo Shan, Shanghai, Nan Tong and Beijing are the top four in our study.

To examine the validity of our analysis, a comparison with official overall CEI is needed. We visualize such comparison by plotting rank difference between our study and official overall CEI for all 284 participating mainland cities from China. The plot can be found in Fig.3. Again, the overall pattern of the data points are much more important than the city names those points correspond to, and we do not label the city names in English. In Figure.3, the horizontal axis represents official CEI values, and vertical axis demonstrates rank difference. The data points are clustered within  $\pm 50$  in rank difference for most of the 284 cities.

$$\text{rank difference} = (\text{rank by official overall CEI}) - (\text{our rank by Z1}). \quad (7)$$

The corresponding bar chart for rank difference is shown in Fig.4. To further examine the consistency between our studies and official CEI results, a paired difference test is performed.

In statistics, a paired difference test is a type of location test that is used when comparing two sets of mea-

surements to assess whether their population means differ. The Wilcoxon signed rank test is a non-parametric statistical hypothesis test, which is used when comparing two related samples, matched samples, or repeated measurements on a single sample to assess whether their population mean ranks differ. In this study, without the assumptions of normality and independence, the Wilcoxon signed rank test is used to test the significance of the differences between each two sets of measurements(i.e. calculated CEI and official CEI).

The results from Wilcoxon signed rank test can be found in Table.V. We tend to accept the null hypothesis. That is to accept the two sets of measurements exhibit no difference.

Tests for Location: Mu0=0			
Test	-Statistics-	- - - -p Value- - - -	
Student's t	t 0	Pr >  t	1.0000
Sign	M 17	Pr ≥  M	0.0484
Signed Rank	S 1699	Pr ≥  S	0.2108

TABLE V. The results from Wilcoxon signed rank test.

#### IV. FURTHER DISCUSSIONS

The Wilcoxon signed rank test indicates a consistency achieved between our studies and official CEI results. Although we only use the first principal component to model our score, the result is already quite telling and explanatory. One possible potential study may be to include more principal components to see if any improvement could be obtained. We defer this topic to a future study.

We already know from Table.III that the cumulative proportion if including first four principal components is up to 82%, which is, under any circumstance, good enough to draw any strong conclusion.

When official China cities CEI is calculated by China CEI research group, a weighted linear summation method is applied to include contributions from various sources. As stated earlier, determining weight is a quite subjective process, although advanced Analytic Hierarchy Process(AHP) method is employed to minimize the effects. The weight for credit launch index is 0.304075, for enterprise risk management it is 0.141675, 0.068975 is determined for construction of credit reporting system, 0.1253 for governmental credit monitoring, 0.1853 for discredit/dishonesty, and 0.082475 for credit education status. Although a weight of 0.0922 is assigned for corporate experience, there is not any categorical CEI data(or categorical city rankings) by this index in CEI-2012, which is the reason why we only include six out of seven first-class indices in our studies.

The achieved consistency also implies those weights assigned by China CEI research group is reasonable.

The readers may have noticed from Fig.3 that there are apparently more cities sitting above “rank difference =

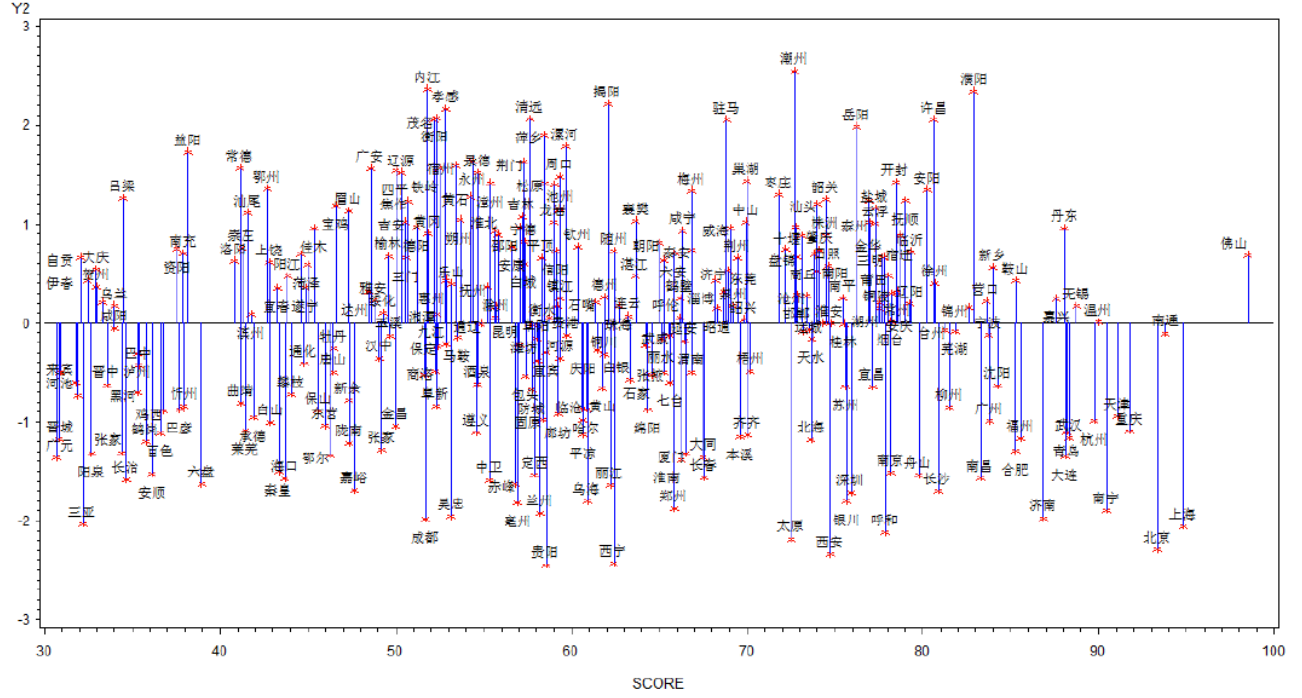


FIG. 2. (City label in Chinese)Score plot for 284 participating China cities. The horizontal axis is a naive mapping of Z1(the first component value) into score( $score = -11 \times Z1 + 61$ ), where Z1 is calculated via Eq.(5). The linear transformation of Z1 is to have our score within 30-100, which is comparable with official CEI that falls into the same range.

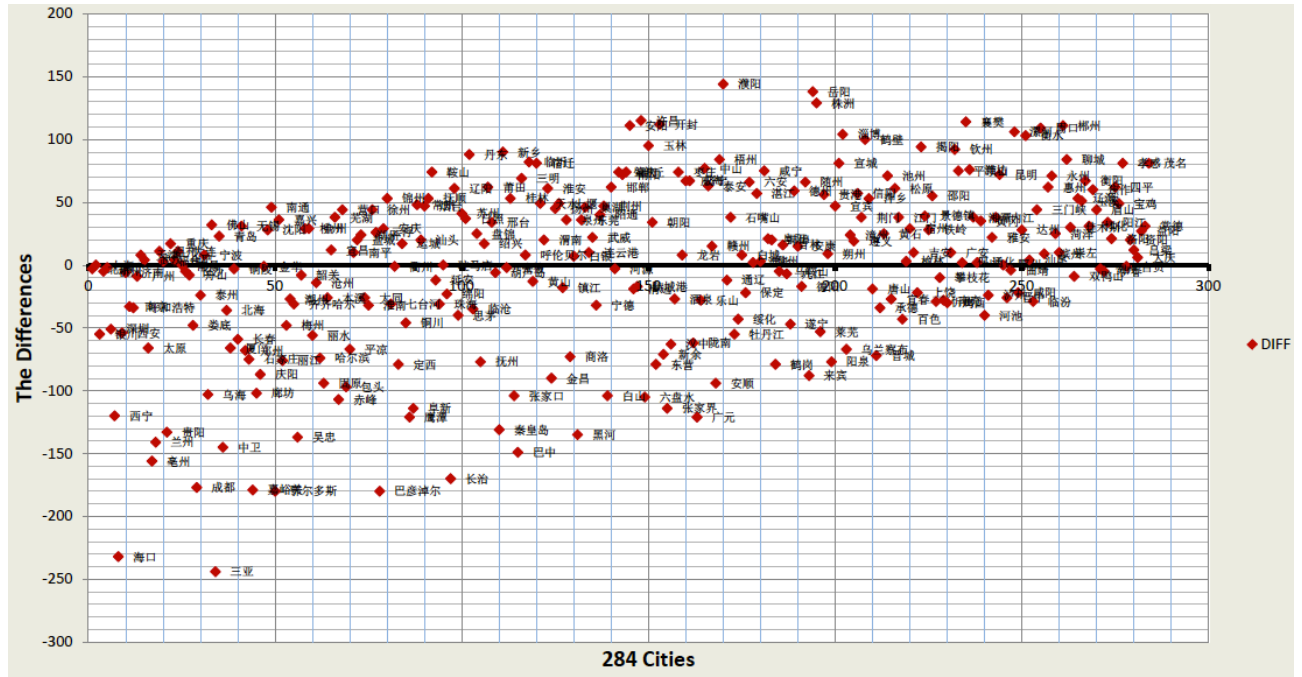


FIG. 3. (City label in Chinese)Rank difference plot for 284 participating China cities. The horizontal axis is official CEI values, and vertical axis is rank difference calculated via Eq.(7)

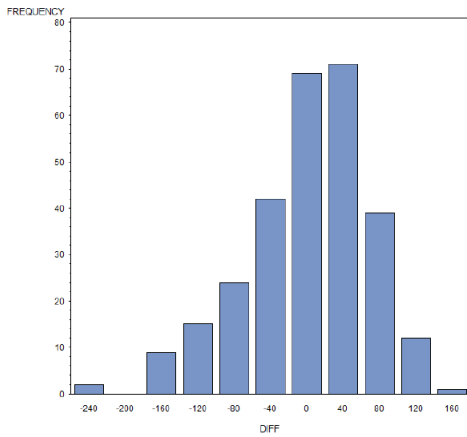


FIG. 4. Frequency plot for rank difference.

0" reference line, which might indicate that official CEI exhibits, if any, an interesting trend to slightly "overestimate" the goodness of China cities commercial credit environment. This interesting feature can also be easily captured from Fig.4, where center of the gaussian-shape distribution is located in the right side of zero. A simple gaussian-like fit shows the distribution is off to right by  $\sim 20$  city ranks. If it is there, this interesting trend would definitely render us a new direction to further adjust our assigned factor weights and/or improve our model in future.

In addition, city rankings vary in the order of  $\sim 30$  in two different scenarios, which can be estimated by the standard deviation of our gaussian fit. This model-dependent feature of rankings indicates that within small enough ranking range (let's say, 30), city rank is less indicative, and thus we cannot draw any clear conclusion

as to whether one city is better/worse than another.

## V. SUMMARY

CEI is designed to reflect various aspects of the comprehensive urban commercial credit environment. For practical reasons, CEI needs to take into account the feasibility of credit information collecting and the authority of data. CEI is constructed based on multiple factors (called index), which are categorized into three major classes. We present an independent statistical study using categorical city rankings by each of the six first-class indices via principal component analysis techniques. We are able to construct, in parallel, an independent CEI that reflects same features of commercial credit environment among all 284 participating mainland cities in China. The consistency between our studies presented in prior section of this paper and official CEI results indicates CEI is indeed a convincing and powerful tool to quantify and evaluate the status quo of commercial environment among mainland cities in China.

Based on our studies, the factor of government-level credit monitoring services and credit education status are the two most influential indices in modeling CEI. The extraction of important factors serves as a guideline for future policy making so as to build a better overall commercial credit environment across major mainland cities in China.

## VI. ACKNOWLEDGEMENT

We would like to thank China CEI research group for giving us the opportunity to participate in China cities CEI studies and providing data for our analysis. We appreciate valuable comments from CEI research group and intriguing discussions with professor Junyue Lin.

- 
- [1] J. Lin, *Principle of social credit system* (China Founder Press, 2003).
  - [2] in Chinese, "The official report for china city commercial environment index," (2010), <http://www.chinacei.org> (official website for China CEI).
  - [3] commercial CEI research group of China, *CEI Blue Book : 2012 Chinese city commercial credit environment index* (2012) pp. 20–21.
  - [4] X. Pu and J. Han, China Business Press, 6 (2011).
  - [5] G. Xu, *Theory and Application of statistical indicators*, Ph.D. thesis, China Statistics Press (2004).
  - [6] J. Lin, *Finance and Trade Economics*, **2**, 14 (2012).
  - [7] G. Bertola, R. Disney, and C. B. Grant, *The economics of consumer credit* (The MIT Press, 2006) pp. 347–358.
  - [8] H. Abdi and L. J. Williams, *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**, 433 (2010).
  - [9] R. Woolson, *Wiley Encyclopedia of Clinical Trials* (1998).